CHE384 Data to Decisions
Chris Mack, University of Texas at Austin

Homework #10 – Design of Experiments – Solutions

Turn in your solution with the answers to the questions below. Also, email to me the supporting spreadsheet and/or R script that you used to perform the analysis. (Please name the file using this format: HW10_yourname.xlsx or HW10_yourname.R).

Consider the full factorial Oxide Thickness data found in the spreadsheet "Design of Experiments.xlsx" using thickness as the response variable.

1. Build a Main Effects model (that is, a model with only linear terms in A, B, C, and D).
    a) Which term (A, B, C, or D) is the most important?

```
Call:  lm(formula = Thickness ~ ., data = oxide.new)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 398.9688     1.4100 282.952  < 2e-16
A            21.4688     1.4100  15.226  < 2e-16
B             9.0313     1.4100   6.405 2.73e-08
C            -5.3125     1.4100  -3.768 0.000383
D            -0.9375     1.4100  -0.665 0.508716

Residual standard error: 11.28 on 59 degrees of freedom
Multiple R-squared:  0.8297, Adjusted R-squared:  0.8182
F-statistic: 71.87 on 4 and 59 DF,  p-value: < 2.2e-16
```

The term with the largest coefficient is most important (since our effects have been normalized). Therefore, A is most important.

    b) Which term(s), if any, are not statistically significant?

D is not statistically significant.

2. Now add all possible interaction terms to the model.
    a) Create an ordered list of terms and their coefficients. Note that the extreme coefficients (very large magnitude) are the most significant ones.

```
Call:  lm(formula = Thickness ~ .^4, data = oxide.new)

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 398.96875    0.23799 1676.389  < 2e-16
A            21.46875    0.23799   90.208  < 2e-16
B             9.03125    0.23799   37.948  < 2e-16
C            -5.31250    0.23799  -22.322  < 2e-16
D            -0.93750    0.23799   -3.939 0.000264
A:B           8.59375    0.23799   36.109  < 2e-16
A:C          -5.62500    0.23799  -23.635  < 2e-16
A:D           0.62500    0.23799    2.626 0.011555
B:C           1.93750    0.23799    8.141 1.35e-10
B:D          -1.68750    0.23799   -7.091 5.34e-09
```

```
C:D           0.34375    0.23799    1.444 0.155131
A:B:C         0.06250    0.23799    0.263 0.793973
A:B:D         1.31250    0.23799    5.515 1.37e-06
A:C:D         0.09375    0.23799    0.394 0.695385
B:C:D         0.15625    0.23799    0.657 0.514617
A:B:C:D       0.46875    0.23799    1.970 0.054670

Residual standard error: 1.904 on 48 degrees of freedom
Multiple R-squared:  0.9961, Adjusted R-squared:  0.9948
F-statistic: 807.6 on 15 and 48 DF,  p-value: < 2.2e-16
```

b) How might you decide which terms are large enough to include in the final model?

Use the p-values: we could decide to include all main effects, and all two-factor interactions except C:D. We might also include one three factor interaction, A:B:D

3. Suppose you decide to include only the terms A, B, C, AB, and AC.
   a) Are all the model terms significant?

```
Call: lm(formula = Thickness ~ A + B + C + A:B + A:C, data = oxide.new)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 398.9688     0.4677  853.03  < 2e-16
A            21.4688     0.4677   45.90  < 2e-16
B             9.0313     0.4677   19.31  < 2e-16
C            -5.3125     0.4677  -11.36 2.24e-16
A:B           8.5937     0.4677   18.37  < 2e-16
A:C          -5.6250     0.4677  -12.03  < 2e-16

Residual standard error: 3.742 on 58 degrees of freedom
Multiple R-squared:  0.9816, Adjusted R-squared:   0.98
F-statistic: 618.2 on 5 and 58 DF,  p-value: < 2.2e-16
```

All of the terms are statistically significant.

b) Compare the model coefficients for this model to the corresponding coefficients from the models in problems 1 and 2. What can you conclude?

The data are orthogonal. Model coefficients don't change if terms are added or removed from the model.